

联合生成对抗网络和检测网络的 SAR 图像目标检测*

韩子硕,王春平,付强,赵斌

(陆军工程大学石家庄校区 电子与光学工程系,河北石家庄 050003)

摘要:针对合成孔径雷达图像目标检测中存在的样本获取困难且数量有限问题,提出了联合生成对抗网络和检测网络的学习模型。利用原始训练集对特别设计的超快区域卷积神经网络进行预训练;通过基于注意力机制的深度学习生成对抗网络生成高质量合成样本,并输入检测网络进行预测;依据预测信息和概率等价类属标签分配策略为新生样本提供注释信息,并以一定占比对原始训练集进行扩充;利用扩充数据集对检测网络进行再训练。多组仿真实验证明,所提框架能够有效提升网络检测效率和性能。

关键词:生成对抗网络;快速区域卷积神经网络;合成孔径雷达;目标检测

中图分类号:TN391 **文献标志码:**A **文章编号:**1001-2486(2022)03-164-12

Target detection in SAR images based on joint generative adversarial network and detection network

HAN Zishuo, WANG Chunping, FU Qiang, ZHAO Bin

(Department of Electronic and Optical Engineering, Shijiazhuang Campus of Army Engineering University, Shijiazhuang 050003, China)

Abstract: Aiming at the problem of difficult and limited sample acquisition in SAR (synthetic aperture radar) image target detection, a learning model of joint GAN (generative adversarial network) and detection network was proposed. The original training set was used to pretrain the specially designed faster regional convolutional neural network. The deep convolutional GAN based on attention mechanism was employed to generate extensive synthetic samples, which were input into the detection network for prediction. The corresponding annotation information of the new samples was determined by the prediction information and probability equivalent class label allocation strategy, and the annotated new samples were used to expand the original dataset with a certain proportion. The detection network was retrained with the expanded dataset. Simulation results show that the proposed framework can improve the detection efficiency and performance of the network effectively.

Keywords: generative adversarial network; faster region-based convolutional neural network; synthetic aperture radar; target detection

合成孔径雷达 (synthetic aperture radar, SAR) 是一种主动式的微波成像传感器,可以全天候、全天时地执行遥感监视任务^[1],其多极化和高分辨率成像模式为对地监测提供了大量可利用数据。以此为基础,SAR 图像目标检测技术得以蓬勃发展,并广泛应用于军事和民用领域。然而,由于 SAR 图像复杂的内在特性和较差的观测效果,传统的目标检测算法^[2-5]收效甚微。随着深度学习算法的崛起,启发于神经网络在可见光图像目标检测领域的巨大成功,研究人员开始将其应用于 SAR 图像目标检测当中,并取得了一定成果,例如车辆检测^[6]、舰船检测^[7-8]、地质检测^[9]等。

深度学习在可见光图像目标检测领域所取得的成就,得益于大数据为其提供源源不断的训练

样本,而相较于可见光图像数量多、易收集的特点,SAR 图像因获取方式单一、代价大,目前公开数据却很少。除此之外,执行数据注释也是一项非常耗费精力和成本的任务。因此,研究人员开始追求以有限数据获取高效网络模型的方法,进而数据增强技术应运而生。目前应用较多的数据增强技术可分为三类:仿射变换^[10-11]、数据合成^[12-13]和数据生成^[14-15]。仿射变换通过图像旋转、拉伸、剪切、色彩变换、区域擦除等图像处理操作获取扩展图像,但其不能深层次地挖掘目标图像特征,因此目标多样性并没有发生根本变化,甚至当使用大量扩展图像训练深度学习模型时,网络将产生过拟合现象^[16]。数据合成通过将目标切片随机镶嵌于新的背景当中获取扩展图像,但其并不能保证合成图像能够始终精确地模拟原始

* 收稿日期:2020-08-25

基金项目:国家部委基金资助项目(LJ20191A040155)

作者简介:韩子硕(1986—),男,河北石家庄人,博士研究生,E-mail:shuo1986andy@126.com;

王春平(通信作者),男,教授,博士,博士生导师,E-mail:chunpw_tom@163.com

图像的视觉环境,而且此方法还需要大量的外部数据支持^[17]。数据生成主要是通过生成对抗网络(generative adversarial network, GAN)来生成新的图像,其强大的特性以及逼真的新生图像,深受业界人士追捧。

GAN由Goodfellow等^[18]于2014年提出,并应用于视觉图像生成,从此许多基于GAN的衍生模型相继而来。深度卷积生成对抗网络^[19](deep convolutional GAN, DCGAN)用卷积神经网络(convolutional neural network, CNN)和全局池化层分别代替GAN中的多层感知机和全连接层,以稳定GAN的训练过程。Arjovsky等^[20]提出的WGAN,通过更加科学地度量生成数据与真实数据之间的差异,很大程度上缓解了梯度消失和模式坍塌问题。Andrew等^[21]提出的BigGAN,通过扩大网络批处理数量和增加卷积通道,并以截断技巧平稳训练过程,使GAN性能得到大幅提升。Zhang等^[22]将注意力机制引入GAN中,使网络能够自我学习应该关注的区域,像素之间的空间关联性也得以增强,解决了卷积结构所带来的感受野受限问题。Karnear等^[23]提出的MSG-GAN,允许梯度流从鉴别器到生成器在多个尺度上流动,为高分辨率图像生成提供了一种稳定的训练方法。随着GAN模型的不断优化,生成图像愈加逼真,其常被应用于各种预测任务,如:文本检测^[24]、字体识别^[25]、目标检测^[26]、人体姿势估计^[27]等。Gaidon等^[28]指出用生成数据对神经网络进行预训练,可提升网络性能。

基于以上分析,本文提出将GAN应用于SAR图像目标检测中,并将其与超快区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)相结合,构建集数据增强与目标检测于一体的深度学习模型,目的是在数

据样本有限的前提下,高效完成目标检测任务。其中,提出了基于自我关注机制的深度卷积生成对抗网络(self-attention DCGAN, SA-DCGAN),生成高质量的数据样本,并利用概率等价类属标签分配策略充分挖掘生成的无标签图像潜力,进而对原始训练集进行数据扩充,最后通过特别设计的Faster R-CNN实现SAR图像目标检测。

1 整体网络框架

为了更加直观地展现本文用于SAR图像的半监督目标检测算法,图1给出了所提算法的基本网络框架。训练过程如下:首先,利用原始训练集对特别设计的Faster R-CNN检测网络进行训练;然后,利用SA-DCGAN生成新的图像样本,并输入检测网络中进行预测,为其生成一个含类别概率和目标位置信息的标签文件;最后,用生成图像扩充训练集,并对检测网络进行微调。下面详细介绍整个过程的具体实现。

1.1 SA-DCGAN

GAN主要包括生成器和鉴别器两个模块。生成器以随机噪声或向量为输入来生成新的图像。鉴别器用于判断输入图像的真假。训练过程中,生成器的目标是尽量生成逼近真实的图像去欺骗鉴别器,鉴别器的目标则是尽量将生成图像与真实图像区别开来,双方相互博弈。在理想状态下,生成器足以生成以假乱真的图像。随着CNN所展现出的强大优势,许多基于GAN的图像生成模型,均利用卷积层来构建生成器和鉴别器,以解决GAN训练不稳定的问题。因此,本文设计了所需的DCGAN模型,并在每一卷积层后引入自我关注模块进行辅助,称之为SA-DCGAN。

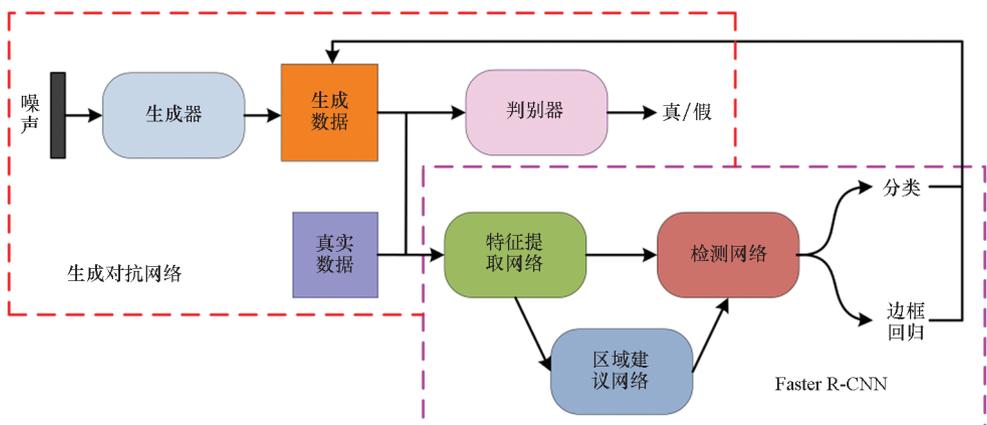


图1 算法整体框架

Fig.1 Overall framework of algorithm

1.1.1 DCGAN

DCGAN 采用 CNN 代替原始 GAN 中的多层感知机,并去掉了 CNN 中的池化层,另外以全局池化层代替全连接层,有力解决了训练不稳定、生成过程不可控、解释性差等缺点。图 2 展示了本文所用 DCGAN 的基本架构。生成器以一个 100 维的随机向量为输入,并利用线性变换将其放大至 $4 \times 4 \times 1\ 024$ 大小,然后通过 5 个核大小为 5×5 、步幅为 2 的反卷积操作,将其扩展为 $128 \times 128 \times 3$ 的图像矩阵。生成器输出层使用 Tanh 激活函数,其余层使用 ReLU 激活函数。与生成器结构相反,鉴别器通过 5 个卷积操作提取训练集和假图像特征,经 sigmoid 函数后,输出判定结果,其所有层均使用 LeakyReLU 激活函数。值得注意的是,除生成器的输出层和鉴别器的输入层外,其他层之后均进行批归一化处理,以确保网络有足够的梯度。

1.1.2 自我关注模型

DCGAN 虽在图像生成领域取得了重大进步,但受 CNN 卷积核限制,其感受野较小,很难表述不同图像区域之间的依赖性,且无法对含有较多几何或结构约束的类别进行建模,而自我关注模型可有效捕捉图像全局依赖性^[29]。自我关注生成对抗网络^[22]在可见光图像生成中,取得了长足进步,但将其应用于 SAR 图像生成却效果欠佳。因此,本文将自我关注模型引入 DCGAN 每层卷积操作之后,构建 SA-DCGAN 模型,提升网络全局建模能力。自我关注模型如图 3 所示,其实现过程如下:

输入:特征图 $u \in \mathbb{R}^{C \times N}$,其中 C 表示特征图通道数, N 表示特征点位置数。

步骤 1:将 u 通过两个 1×1 卷积转换为特征空间 f 和 $g, f(u) = W_f u, g(u) = W_g u$,其中 $W_f \in \mathbb{R}^{\tilde{C} \times C}$ 和 $W_g \in \mathbb{R}^{\tilde{C} \times C}$ 表示 1×1 卷积的权重矩阵, $\tilde{C} = C/8$ 表示卷积通道数。

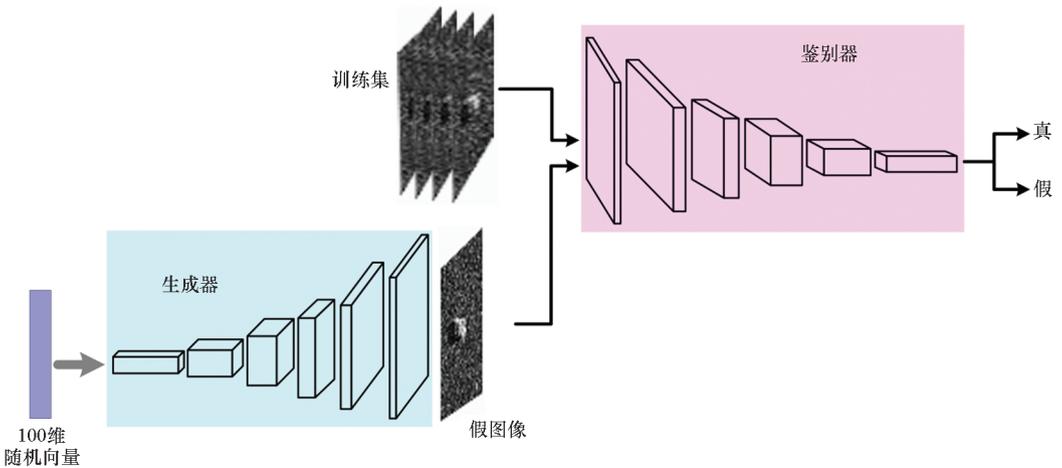


图 2 DCGAN 基本架构
Fig.2 Framework of DCGAN

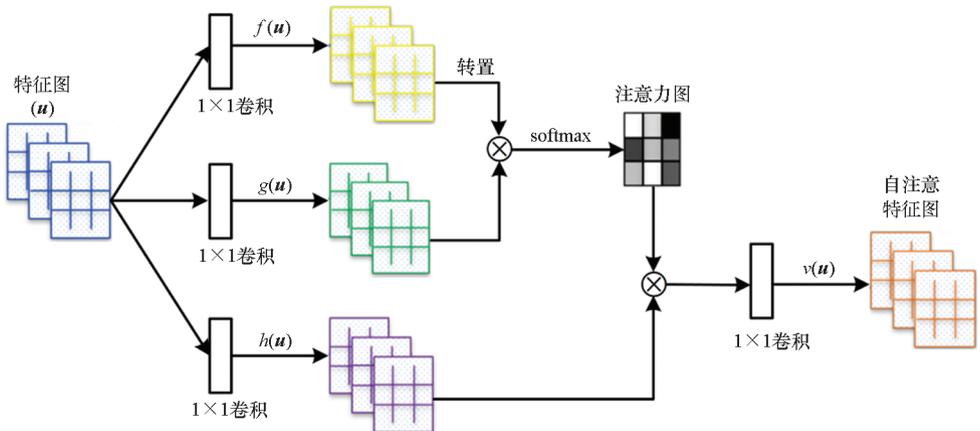


图 3 SA-DCGAN 中引入的自我关注模型
Fig.3 Self-attention model introduced in SA-DCGAN

步骤2:将 f 的输出 $f_{u_i} = f(u_i)$ 经转置后与 g 的输出 $g_{u_j} = g(u_j)$ 相乘构建初始注意力图 $\xi_{j,i} = \exp(s_{ij}) / \sum_{i=1}^N \exp(s_{ij})$,其中 $s_{ij} = f_{u_i}^T g_{u_j}$, $\xi_{j,i}$ 表示模型在合成第 j 个像素时对第 i 个像素的关注度。初始注意力图描述的是原始特征图中一个像素位置与其他像素位置的关系。

步骤3:计算自我关注特征图 $o = (o_1, o_2, \dots, o_j, \dots, o_N) \in \mathbb{R}^{C \times N}$,其中 $o_j = v[\sum_{i=1}^N \xi_{j,i} h(u_i)]$, $h(u_i) = W_h u_i, v(u_i) = W_v u_i, W_h \in \mathbb{R}^{\tilde{c} \times c}$ 和 $W_v \in \mathbb{R}^{\tilde{c} \times c}$ 同样表示 1×1 卷积的权重矩阵。

步骤4:将自我关注特征图 o 乘以占比参数 γ 后,与原始特征图 u 相加,其结果作为自我关注模块的最终输出,可表示为: $\zeta_i = \gamma o_i + u_i$,其中 γ 为可学习参数,初始化为0。

1.1.3 SA-DCGAN 训练设置

在训练之前,首先将训练图像转换为 128×128 大小,并映射至 $[-1, 1]$ 范围。所有卷积

层初始化参数均从均值为0、标准差为0.02的正态分布中随机采样得到,批量大小设置为128。LeakyReLU激活函数中 α 取值为0.2。训练过程中,采用Adam优化器^[30]($\beta_1 = 0, \beta_2 = 0.9$)加速网络训练,并设置生成器学习率为0.0001、鉴别器学习率为0.0004,两者相互交替训练。网络生成对抗损失函数采用Hinge-loss^[31]:

$$L_D = -\mathbb{E}_{(q,d) \sim p_{\text{data}}} \{ \min[0, -1 + D(q, d)] \} - \mathbb{E}_{z \sim p_z, d \sim p_{\text{data}}} \{ \min[0, -1 - D(G(z), d)] \} \quad (1)$$

$$L_G = -\mathbb{E}_{z \sim p_z, d \sim p_{\text{data}}} D(G(z), d) \quad (2)$$

其中, \mathbb{E} 表示求取期望, $D(\cdot)$ 表示鉴别器输出, $G(\cdot)$ 表示生成器输出, q 和 d 为从真实分布 p_{data} 中采样得到的样本, z 为从随机分布 p_z 中采样得到的样本。

1.2 基于Faster R-CNN的目标检测网络

Faster R-CNN模型共包含三个模块:特征提取网络、区域建议网络(region proposal network, RPN)和检测网络,具体框架结构见图4。

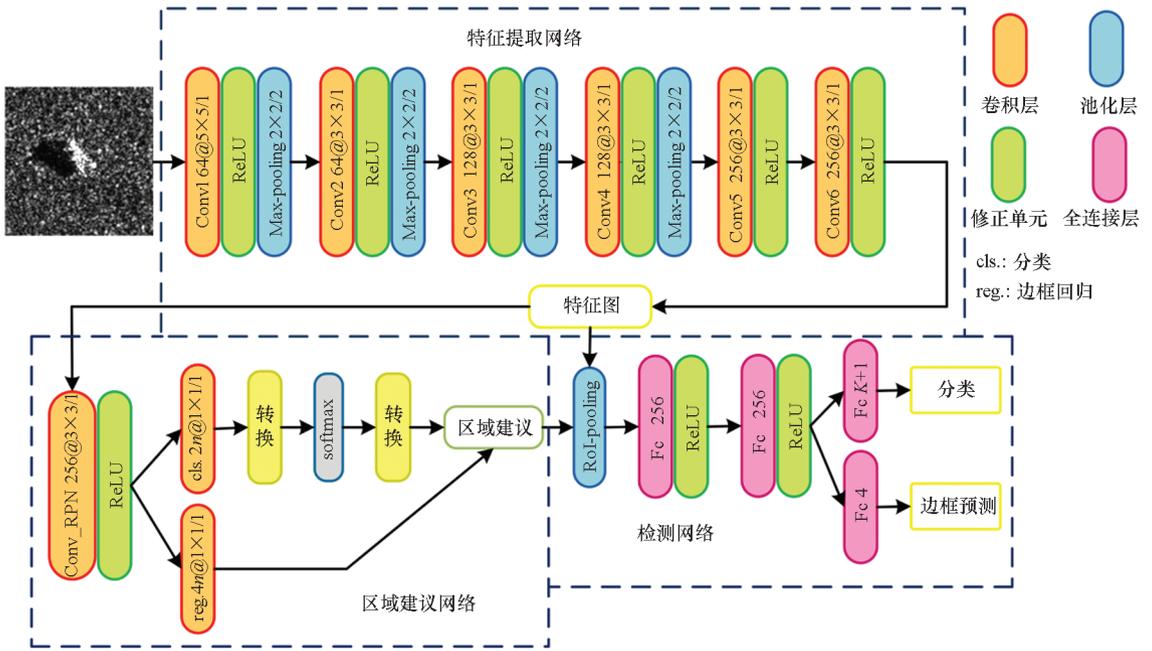


图4 Faster R-CNN 框架

Fig.4 Framework of Faster R-CNN

1.2.1 特征提取网络

特征提取网络旨在获取输入图像特征,为RPN和检测网络提供表征能力强大的特征图。本文设计的特征提取网络共有6层(Conv1~Conv6),包括6个卷积层,除第一层卷积核大小为5外,其余层卷积核大小均为3,卷积核滑动步长取1,卷积核个数依次为64、64、128、128、256、

256。在前四层卷积之后紧接着最大池化层Max-pooling,池化层大小和滑动步长均为2。所有层都使用ReLU作为激活函数,并且在ReLU之后执行批归一化。

1.2.2 区域建议网络

RPN作为整个框架的重要组件,通过合理的锚框设置来快速完成高质量候选区域提取,为检测网络提供合理化建议。设置锚框比例为

$\{1:1, 1:2, 1:3, 1:5, 5:1, 3:1, 2:1\}$, 尺度取 $\{10^2, 16^2, 32^2, 48^2, 64^2\}$ 。在特征提取网络输出的特征图上,以滑窗操作生成多个锚框,经一个 3×3 卷积层后,通过两个全连接层为每个锚框分配一个二进制类标签(是/不是目标)和四个参数坐标 (x, y, w, h) , 其中 (x, y) 表示 RPN 预测的边框中心点坐标, w 和 h 分别表示宽与高,具体过程如图 5 所示。

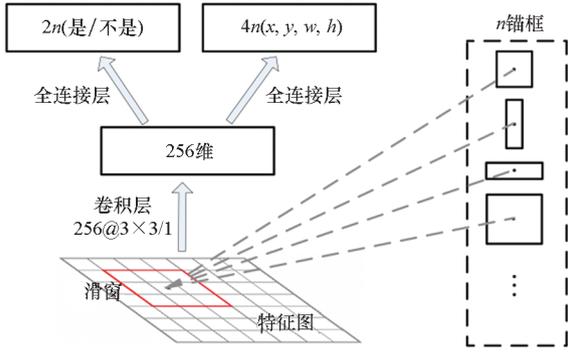


图 5 RPN 示意

Fig. 5 RPN schematic diagram

在 RPN 训练阶段,为了提升计算效率,需要将所有的锚框分为正样本和负样本。正样本定义为:与任意地面真值交并比(intersection-over-union, IoU)大于 0.7 的锚框。负样本定义为:与所有地面真值 IoU 均小于 0.3 的锚框,其余非正非负的样本被舍弃。从正负样本中随机抽取 256 个作为一个小批量训练集,正负样本保持 1:1 比例,如果正样本个数小于 128,则用负样本填补。RPN 网络初始学习率为 0.001,每 20 000 次迭代将当前的学习率除以 10,初始预训练最大迭代次数为 40 000,二次训练最大迭代次数为 20 000。通过 RPN 提取到的候选区域之间会有大量的重叠,基于候选区域分类得分,采用非极大值抑制(non-maximum suppression, NMS)去除重叠的候选区域,并将得分最高的前 200 个候选区域作为感兴趣区域(region of interests, RoIs)输入检测网络中。

1.2.3 检测网络

检测网络用于执行对候选区域的最终判定,其输入为由特征提取网络构建的特征图和由 RPN 提供的候选区域。首先,计算每个候选区域在特征图上对应的映射区域,然后利用 RoI 池化将这些映射区域统一到相同的大小,经两个全连接层(fully convolutional network, FCN)后,由分类层输出每个候选区域所属类别的概率分布 $\mathbf{p}^* = (p_0^*, p_1^*, \dots, p_K^*)$, 其中 K 表示类别

数,根据 \mathbf{p}^* 为每个候选区域添加类别标签,同时由边框回归层输出目标的参数化坐标,指定与地面真值相交不小于 0.5 的候选区域为正样本。

检测网络训练参数设置方式与 RPN 一致,初始预训练最大迭代次数为 60 000,二次训练最大迭代次数为 30 000。检测网络和 RPN 共享特征提取网络,并采用交替方式进行训练^[32]。

1.2.4 概率等价类属标签分配

目前,为数据增强生成的无标签图像添加类属标签共有三种处理方法:一是新增类处理^[33],即在原始数据集上新增加一个类别,将所有新生样本均归为这一类别;二是最大概率类属标签分配^[34],即根据新生样本在网络中的预测结果,将其归为预测可能性最大的类别;三是标签平滑正则化处理^[26],即将新生样本所属类别概率平均分配至所有类,假设原数据集共有 K 类,那么新生样本属于第 i 类的概率为 $1/K$ 。另外,对于新生样本中目标位置信息,均采用网络预测的方式进行标定。

以上方法虽可以在不同程度上提升网络性能,但均不能有效挖掘生成图像潜力。因此,本文提出了概率等价类属标签分配策略,如图 6 所示,其具体过程如下:①利用原始训练集对 Faster R-CNN 进行初始训练;②将生成图像输入检测网络进行预测,并记录预测目标的坐标信息和所属类别概率分布 $\mathbf{p}^* = (p_0^*, p_1^*, \dots, p_K^*)$;③将坐标信息和所属类别概率分布写入注释文件,作为该生成图像的注释。

1.2.5 损失函数

Faster R-CNN 需完成目标分类和定位两项任务,因此使用多任务损失函数,包括分类损失(class loss)和边框回归损失(bbox regression loss)两部分,定义为:

$$L(p_i, l_i, \mathbf{t}_i^*, \mathbf{t}_i) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, l_i) + \lambda \frac{1}{N_{\text{reg}}} \sum_i l_i L_{\text{reg}}(\mathbf{t}_i^*, \mathbf{t}_i) \quad (3)$$

式中, λ 表示平衡系数,用于控制分类损失函数 L_{cls} 与边框回归损失函数 L_{reg} 所占比重, N_{cls} 表示 mini-batch 中输入样本的个数(本文设置 $N_{\text{cls}} = 256$), N_{reg} 表示特征图大小, p_i 表示样本预测为目标概率, i 表示一个 mini-batch 中样本的索引, l_i 表示第 i 个样本的真实标签, \mathbf{t}^i 表示定位边框的 4 个参数化坐标, \mathbf{t}_i^* 表示真值区域的 4 个参数化坐标。

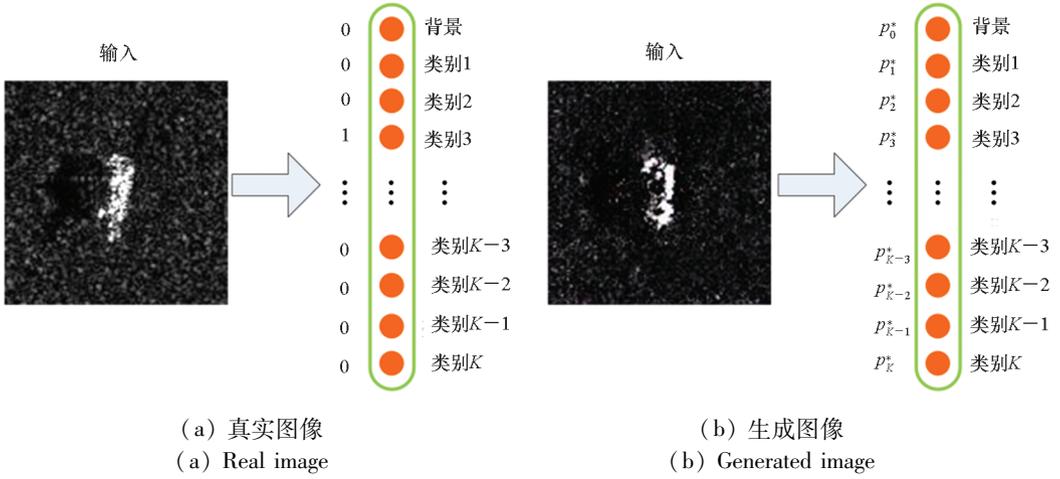


图6 真实图像与生成图像的分类标签分布

Fig. 6 Distribution of the labels of real images and generated images

式(3)中 L_{cls} 和 L_{reg} 分别定义为:

$$L_{\text{cls}}(p, l) = - (1 - \eta) \ln p^l - \eta \sum_{k=1}^K p_k^* \ln(p(k))^l \quad (4)$$

$$L_{\text{reg}}(\mathbf{t}^*, \mathbf{t}) = (1 - \eta) s_{L_1}(\mathbf{t}^* - \mathbf{t}) + \varepsilon \eta s_{L_1}(\mathbf{t}^* - \mathbf{t}) \quad (5)$$

其中:当第 i 个样本取自真实图像时, $\eta = 0$; 取自生成图像时, $\eta = 1$ 。 ε 用于调整生成图像边框回归损失占比, $s_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases}$ 。 $\mathbf{t} = (t_x, t_y, t_w, t_h)$ 与 $\mathbf{t}^* = (t_x^*, t_y^*, t_w^*, t_h^*)$ 的具体形式为:

$$\begin{cases} t_x = (x - x_a) / w_a \\ t_y = (y - y_a) / h_a \\ t_w = \ln(w / w_a) \\ t_h = \ln(h / h_a) \end{cases} \quad (6)$$

$$\begin{cases} t_x^* = (x^* - x_a) / w_a \\ t_y^* = (y^* - y_a) / h_a \\ t_w^* = \ln(w^* / w_a) \\ t_h^* = \ln(h^* / h_a) \end{cases} \quad (7)$$

式中: x, x_a 和 x^* 分别表示预测框、锚框和标记框的中心点横坐标; y, w, h 的定义方式同上。

2 实验设置及评估

所有实验均在配置为 E5 - 2630v4 CPU、NVIDIA GTX - 1080Ti GPU (11 G video memory)、64 GB RAM 的图形工作站上进行,以深度学习框架 TensorFlow^[35] 为编译工具完成。

2.1 数据集

为了验证本文算法的有效性,在实验中使用

了移动和静止目标获取与识别 (moving and stationary target acquisition and recognition, MSTAR) 数据集和 SAR 舰船检测数据集 (SAR ship detection dataset, SSDD)。

MSTAR 数据集由美国国防高级研究计划局和空间实验室发布。该数据集由 X 波段、HH 极化方式、0.3 m × 0.3 m 高分辨率聚束式合成孔径雷达采集得到,共包括 10 类典型军事目标在 0 ~ 360° 不同方位角下的静止切片图像。实验过程中,使用 BMP2、BTR70、T72 图像切片作为实验对象,图 7 所示为这三类军事目标在真实场景下的可见光图像及其 SAR 图像。数据集详细信息见表 1。



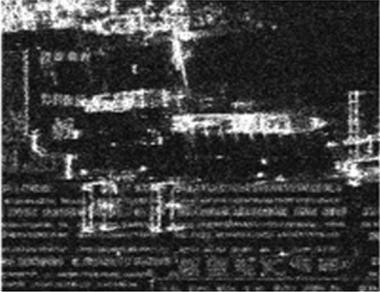
图7 BMP2、BTR70 和 T72 的 SAR 图像和光学图像

Fig. 7 SAR image and optical image of BMP2, BTR70 and T72

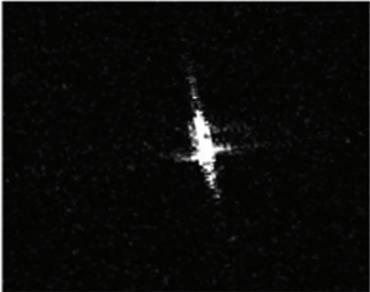
表1 MSTAR 数据集详细信息
Tab. 1 Details of MSTAR dataset

类别	BMP2	BTR70	T72
训练集	233 (Sn_9563)	233	232 (Sn_132)
测试集	196 (Sn_9566)	196	195 (Sn_812)

SSDD 数据集是由 Li 等^[36] 构建的 SAR 图像舰船检测数据集,共包含 2 456 艘舰船的 1 160 张图像切片。该数据集中的图像切片收集于 RadarSat - 2、TerraSAR - X 和 Sentinel - 1。数据集中图像分辨率和尺寸均不固定,样例见图 8,其详细信息见表 2。



(a) 高分辨率图像
(a) High resolution image



(b) 中分辨率图像
(b) Medium resolution image



(c) 低分辨率图像
(c) Low resolution image

图 8 SSDD 数据集样例
Fig. 8 Samples of SSDD dataset

表 2 SSDD 数据集详细信息
Tab. 2 Details of SSDD dataset

类别	图像切片	舰船数
训练集	800	1 827
测试集	360	629

2.2 评估参数

为定量分析 SA-DCGAN 性能,使用弗雷切特

初始距离^[37] (Frechet inception distance, FID) 和最大均值差异^[38] (maximum mean discrepancy, MMD) 两项指标对模型进行评估。FID 计算的是真实图像和生成图像在特征层面的距离, MMD 用于判断两个分布之间的相似程度。FID 与 MMD 的值越小说明真实图像与生成图像越相似。

为了验证网络检测性能,使用平均精度 (average precision, AP) 和 mAP 两项指标对其进行评估。这两项指标均由真阳性 (true positive, TP)、假阳性 (false positive, FP)、假阴性 (false negative, FN)、真阴性 (true negative, TN) 四个分量计算获得。TP 和 FP 分别表示正确预测和错误预测的数量, FN 表示漏检数量。精确率 P 和召回率 R 定义为:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

AP 可反映网络对单个类别的检测性能,是检测领域常用的一种统计量,其值越大说明网络性能越好。

$$AP = \int_0^1 P(R) dR \quad (10)$$

mAP 是数据集所有类 AP 的均值,用于评估网络的全局性能。

3 实验结果及分析

3.1 SA-DCGAN 性能评估

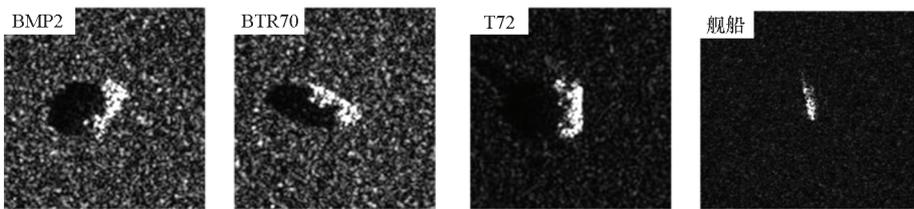
本文通过将 SA-DCGAN 与 SAGAN^[22]、DCGAN^[19] 进行比较来验证所提生成对抗网络模型的优越性。三种方法的定量分析见表 3,生成图像示例见图 9。通过表 3 可以看出,无论是 MSTAR 数据集还是 SSDD 数据集,SA-DCGAN 生成图像的 FID 和 MMD 都明显低于其他两种方法,说明本文设计的生成对抗网络模型效果最优。MSTAR 数据集图像包含目标、阴影、背景三个要素,而 SSDD 数据集图像包含目标、船尾迹、背景三个要素。由图 9 可知,SAGAN 生成图像的目标几何外观发生严重畸变,但对各数据集中三要素刻画比较成功,三要素对比度强于 DCGAN。DCGAN 生成图像虽在外观上较 SAGAN 有良好改善,但三要素边界混淆严重,甚至船尾迹发生丢失。SA-DCGAN 生成图像的目标外观和三要素边界均与真实图像逼近,甚至 T72 炮管的阴影和船尾迹呈现均清晰可辨。

GAN 可以产生逼真的样例,但训练过程却十分困难,很容易出现不收敛的情况。良好的 GAN 训练态势,需要确保鉴别器损失不能一直下降,而

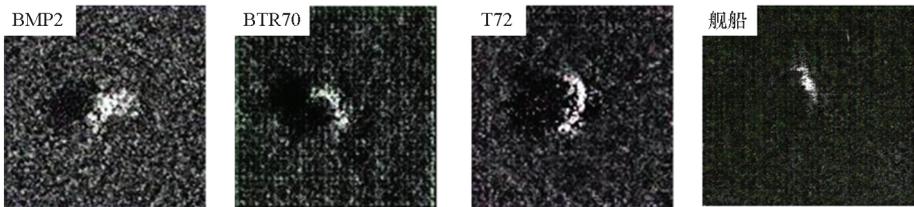
生成器损失不能一直上升,理想情况下,两者最终都应在某一数值附近波动。图 10 展示了 SA-DCGAN 在 MSTAR 数据集和 SSDD 数据集上的损

表 3 三种 GAN 的性能对比
Tab. 3 Performance comparison of three GANs

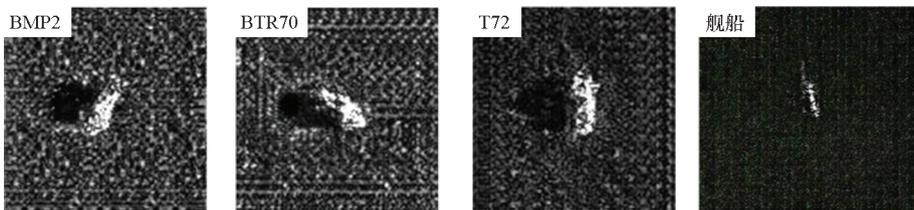
方法	MSTAR 数据集						SSDD 数据集	
	BMP2		BTR70		T72		FID	MMD
	FID	MMD	FID	MMD	FID	MMD		
SAGAN	2.355	0.568	1.963	0.731	2.077	0.594	2.872	0.737
DCGAN	1.576	0.475	1.850	0.694	1.467	0.528	2.508	0.632
SA-DCGAN	1.434	0.388	1.649	0.412	1.344	0.307	2.156	0.516



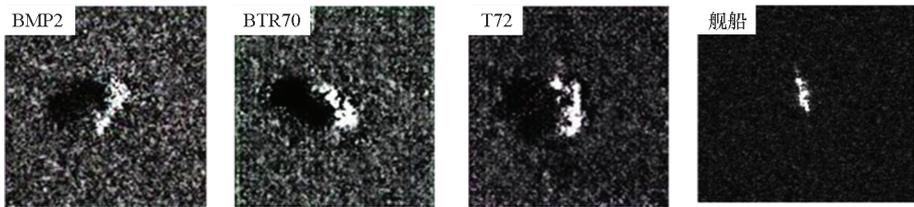
(a) 真实图像
(a) Real images



(b) DCGAN 生成图像
(b) Images generated by DCGAN



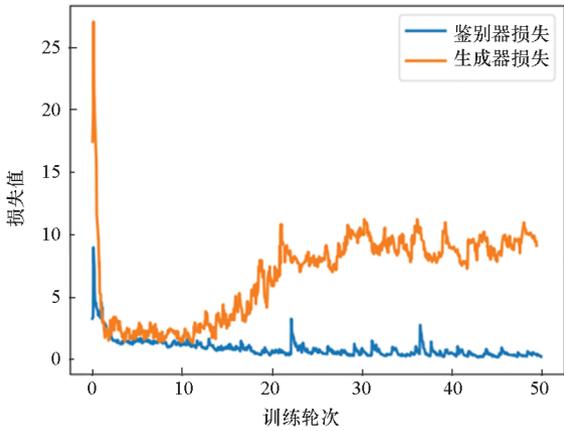
(c) SAGAN 生成图像
(c) Images generated by SAGAN



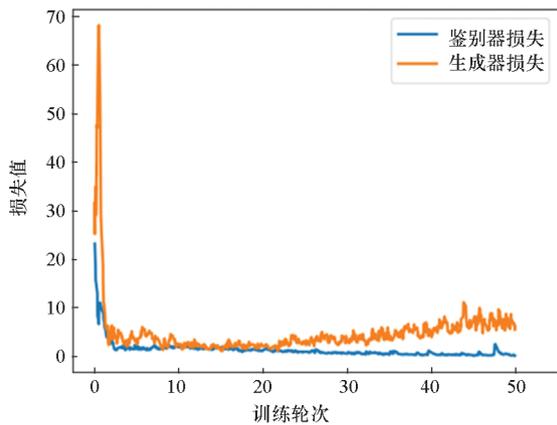
(d) SA-DCGAN 生成图像
(d) Images generated by SA-DCGAN

图 9 生成图像示例

Fig. 9 Examples of generated images



(a) MSTAR 数据集上的损失曲线
(a) Loss on MSTAR dataset



(b) SSDD 数据集上的损失曲线
(b) Loss on SSDD dataset

图 10 SA-DCGAN 的损失曲线
Fig. 10 Loss of SA-DCGAN

鉴别器损失。由图可知,损失曲线在两个数据集上的走势基本一致,随着训练轮次的增加,最终生成器损失逐渐收敛于 8 附近,鉴别器损失逐渐收敛于 0.6 附近,说明 SA-DCGAN 训练情况理想。虽然良好的收敛性并不能证明网络的性能,但多个数据集上相同的损失曲线收敛态势可以侧面说明所提模型具有优异的普适性和鲁棒性。

3.2 训练集中生成图像占比对检测性能的影响

本小节在 MSTAR 数据集和 SSDD 数据集上分别进行四组实验来验证使用生成样本扩充训练集对网络检测性能的影响。各实验在原始训练样本的基础上扩充不同数量占比的生成图像作为训练集,以所提网络框架为训练和检测模型,实验设置和参数均保持一致,根据最终检测结果确定最优生成样本占比。四组实验的训练集中原始图像与生成图像比例分别为 1 : 0, 2 : 1, 1 : 1, 2 : 3。表 4 列出了两个数据集在不同生成图像占比下的网络检测结果。对于 MSTAR 数据集,由于样本相对简单且干扰信号少,因此当训练集中原始图像与生成图像比例为 2 : 1 时,检测结果就可达到最优,而随着生成图像占比继续增加,混淆信息也随之增多,网络性能开始下降。对于 SSDD 数据集,由于目标尺度多样,原始训练集中缺少一些姿态目标,因此当原始图像与生成图像比例为 1 : 1 时,检测网络才达到最优。根据以上实验结果和分析,此后的基于所提网络模型的检测任务, MSTAR 训练集中生成图像占比设置为 1/3, SSDD 数据集中生成图像占比设置为 1/2。

失曲线,其中黄色曲线为生成器损失,蓝色曲线为

表 4 不同生成图像占比下的检测结果

Tab. 4 Detection results under different proportions of generated images

原始图像与 生成图像比例	MSTAR 数据集				SSDD 数据集
	BMP2 (AP)	BTR70 (AP)	T72 (AP)	mAP	mAP
1 : 0	0.970 8	0.971 4	0.976 2	0.972 8	0.873 5
2 : 1	0.996 3	0.999 7	0.999 8	0.998 6	0.899 1
1 : 1	0.990 9	0.998 7	0.999 9	0.996 5	0.936 8
2 : 3	0.965 3	0.961 4	0.968 8	0.965 2	0.914 2

3.3 不同类属标签分配策略对检测性能的影响

本小节分别将最大概率类属标签分配^[34]、新增类处理^[33]、标签平滑正则化处理^[26]和概率等价类属标签分配 4 种策略应用于网络检测模型,来验证所提标签分配策略的优越性。表 5 列出了

所提网络模型应用不同类属标签分配策略处理生成图像时,其在两个数据集上的性能表现。由表可以看出,应用概率等价类属标签分配策略的网络模型在 MSTAR 数据集和 SSDD 数据集上均得到了最优的检测结果。

表 5 不同类属标签分配策略下的检测结果

Tab. 5 Detection results under different generic label assignment strategies

标签分类算法	MSTAR 数据集				SSDD 数据集
	BMP2 (AP)	BTR70 (AP)	T72 (AP)	mAP	mAP
最大概率类属标签分配	0.991 5	0.992 7	0.996 7	0.993 6	0.893 2
新增类处理	0.994 5	0.999 7	0.996 7	0.996 9	0.915 4
标签平滑正则化处理	0.997 2	0.998 0	0.998 5	0.997 9	0.923 6
概率等价类属标签分配	0.996 3	0.999 7	0.999 8	0.998 6	0.936 8

3.4 与其他检测算法比较

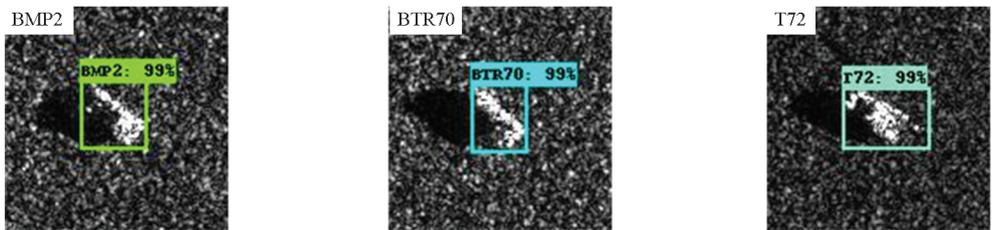
本小节通过将所提网络模型与 Light level CNN^[39]、单发多盒检测器 (single shot multi-box detector, SSD^[40])、改进 Faster R-CNN^[41] 进行比较验证算法的有效性。表 6 列出了不同网络模型的检测结果及平均检测一幅图像 (从输入至输出) 所需的时间,图 11 展示了本文算法在两个数据集上的检测结果示例。Light level CNN 的结构相对简单,包括 2 个卷积层、2 个池化层和 2 个全

连接层,在 MSTAR 数据集和 SSDD 数据集上复现检测结果,mAP 分别达到 0.922 9 和 0.826 3。SSD 为单阶段网络模型,通过在不同尺度的特征层上对目标进行检测来提升网络性能,在两个数据集上的 mAP 分别达到 0.956 9 和 0.870 9。改进 Faster R-CNN 利用初始网络扩展 ResNet101 感受野来提取丰富的图像特征,并利用感兴趣区域对齐单元完成特征池化,其在两个数据集上的 mAP 分别达到 0.984 7 和 0.923 0。相比其他三

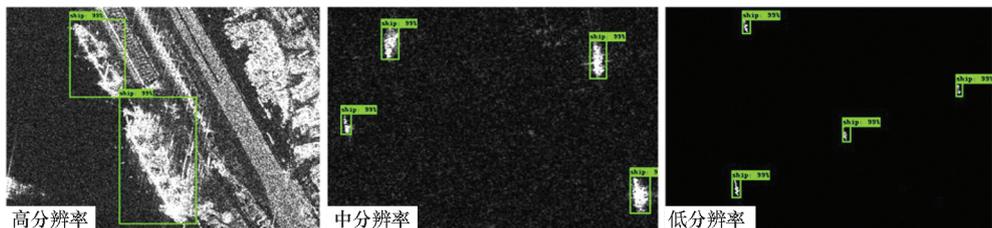
表 6 不同网络模型的检测性能

Tab. 6 Detection performance of different models

网络模型	MSTAR 数据集					SSDD 数据集	
	BMP2 (AP)	BTR70 (AP)	T72 (AP)	mAP	时间/s	mAP	时间/s
Light level CNN	0.923 1	0.918 8	0.926 9	0.922 9	0.37	0.826 3	0.56
SSD	0.951 5	0.959 4	0.959 8	0.956 9	0.23	0.870 9	0.27
改进 Faster R-CNN	0.982 8	0.985 4	0.986 0	0.984 7	0.81	0.923 0	1.04
本文算法	0.996 3	0.999 7	0.999 8	0.998 6	0.41	0.936 8	0.68



(a) MSTAR 数据集检测结果
(a) Detection results of MSTAR dataset



(b) SSDD 数据集检测结果
(b) Detection results of SSDD dataset

图 11 本文算法检测结果示例

Fig. 11 Examples of detection results of the proposed algorithm

种算法,本文算法的检测结果最优。从各网络模型检测一幅测试图像所需平均时间看,SSD 时间损耗最小,Light level CNN 和本文算法时间损耗相近,而改进 Faster R-CNN 时间损耗最大。由以上比较分析可以得出,本文检测模型虽然简单,但在 GAN 的支持辅助下,同样可以获得较优的检测结果,且检测时间损耗远远小于性能相近的其他模型。

4 结论

针对 SAR 图像目标检测中,样本数据获取困难且数量有限的问题,本文提出了一种联合生成对抗网络和检测网络的深度学习模型,可以在有限数据条件下,稳定提升检测性能。所建网络框架首先利用基于注意力机制的深度卷积生成对抗网络合成高质量新生样本,并通过概率等价类属标签分配策略为每个新生样本提供注释信息,再用一定比例的生成图像扩充原始训练集,有效提升 Faster R-CNN 检测性能。多组基于 MSTAR 数据集和 SSDD 数据集的实测实验数据证明,所提 SA-DCGAN 模型可生成逼近真实图像的合成样本,为原始训练集提供扩充数据。概率等价类属标签分配策略无须人工注释,即可为新生样本提供注释信息,满足训练要求,后经定量分析最优生成图像占比,并以此扩充原始训练集,训练特别设计的 Faster R-CNN 检测网络。相较于 Light level CNN、SSD 和改进 Faster R-CNN 模型,本文所提模型可以在较少时间损耗的前提下,高效完成目标检测任务。

所提网络模型虽然取得了优异的检测结果,但其高光表现以大量的训练前准备工作为基础。此外,为有效利用新生样本而提出的概率等价类属标签分配策略,虽无须人工添加注释信息,但其需要原始检测网络具有较高的检测性能。另外,将基于一个数据集训练的模型用于相似数据集检测,其检测结果还不是很理想(如,使用 AIR-SARShip 数据集^[42]对 SSDD 训练的模型进行验证,其 mAP 值达到 0.784 2)。下一步工作,将着重加强三方面的研究:一是进一步融合生成网络和检测网络,简化训练前准备工作;二是进一步探寻有效利用新生样本的方法,令其适用于多种网络模型;三是进一步研究网络模型在相似数据集之间的可推广性。

参考文献 (References)

[1] SHAO J Q, QU C W, LI J W. A performance analysis of

- convolutional neural network models in SAR target recognition[C]// Proceedings of SAR in Big Data Era: Models, Methods and Applications, 2017: 1-6.
- [2] AO W, XU F, LI Y C, et al. Detection and discrimination of ship targets in complex background from spaceborne ALOS-2 SAR images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(2): 536-550.
- [3] WANG C L, BI F K, ZHANG W P, et al. An intensity-space domain CFAR method for ship detection in HR SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(4): 529-533.
- [4] LENG X G, JI K F, FAN Q J, et al. A novel adaptive ship detection method for spaceborne SAR imagery [C]// Proceedings of IEEE International Geoscience and Remote Sensing Symposium, 2016: 108-111.
- [5] ZHAO Z, JI K F, XING X W, et al. Ship surveillance by integration of space-borne SAR and AIS-review of current research[J]. Journal of Navigation, 2014, 67(1): 177-189.
- [6] CUI Z Y, DANG S H, CAO Z J, et al. SAR target recognition in large scene images via region-based convolutional neural networks[J]. Remote Sensing, 2018, 10(5): 776.
- [7] AN Q Z, PAN Z X, YOU H J. Ship detection in Gaofen-3 SAR images based on sea clutter distribution analysis and deep convolutional neural network [J]. Sensors, 2018, 18(2): 334.
- [8] 韩子硕,王春平,付强,等. 基于超密集特征金字塔网络的 SAR 图像舰船检测[J]. 系统工程与电子技术, 2020, 42(10): 2214-2222.
- HAN Z S, WANG C P, FU Q, et al. Ship detection in SAR images based on super dense feature pyramid networks[J]. Systems Engineering and Electronics, 2020, 42(10): 2214-2222. (in Chinese)
- [9] MOHAJERANI S, KRAMMER T A, SAEEDI P. A cloud detection algorithm for remote sensing images using fully convolutional neural networks[C]//Proceedings of IEEE 20th International Workshop on Multimedia Signal Processing, 2018.
- [10] ZHONG Z, ZHENG L, KANG G L, et al. Random erasing data augmentation[EB/OL]. (2017-08-16) [2020-08-20]. <https://arxiv.org/abs/1708.04896v2>.
- [11] CUBUK E D, ZOPH B, MANÉ D, et al. AutoAugment: learning augmentation strategies from data[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 113-123.
- [12] DWIBEDI D, MISRA I, HEBERT M. Cut, paste and learn: surprisingly easy synthesis for instance detection [C]// Proceedings of IEEE International Conference on Computer Vision, 2017: 1310-1319.
- [13] TRIPATHI S, CHANDRA S, AGRAWAL A, et al. Learning to generate synthetic data via compositing[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 461-470.
- [14] CHEN H T, WANG Y H, XU C, et al. Data-free learning of student networks [C]//Proceedings of IEEE/CVF International Conference on Computer Vision, 2019: 3513-3521.
- [15] SHAHAM T R, DEKEL T, MICHAELI T. SinGAN: learning a generative model from a single natural image [C]// Proceedings of IEEE/CVF International Conference on Computer Vision, 2019: 4569-4579.
- [16] PEREZ L, WANG J. The effectiveness of data augmentation in image classification using deep learning[EB/OL]. (2017-

- 12 - 13) [2020 - 08 - 20]. <https://arxiv.org/abs/1712.04621>.
- [17] WANG H, WANG Q L, YANG F, et al. Data augmentation for object detection via progressive and selective instance-switching[EB/OL]. (2019 - 06 - 02) [2020 - 08 - 20]. <http://arxiv.org/abs/1906.00358v2>.
- [18] GOODFELLOW I, JEAN P A, MIRZA M, et al. Generative adversarial nets [C]//Proceedings of 27th International Conference on Neural Information Processing Systems, 2014: 2672 - 2680.
- [19] ALEC R, LUKE M. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. (2015 - 11 - 19) [2020 - 8 - 20]. <https://arxiv.org/abs/1511.06434>.
- [20] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[EB/OL]. (2017 - 01 - 26) [2020 - 08 - 20]. <https://arxiv.org/abs/1701.07875v1>.
- [21] ANDREW B, JEFF D, KAREN S. Large scale GAN training for high fidelity natural image synthesis[EB/OL]. (2018 - 09 - 28) [2020 - 08 - 20]. <https://arxiv.org/abs/1809.11096v1>.
- [22] ZHANG H, GOODFELLOW I J, METAXAS D N, et al. Self-attention generative adversarial networks [EB/OL]. (2018 - 03 - 21) [2020 - 08 - 20]. <http://export.arxiv.org/abs/1805.08318>.
- [23] KARNEWAR A, WANG O. MSG-GAN: multi-scale gradients for generative adversarial networks [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7796 - 7805.
- [24] GUPTA A, VEDALDI A, ZISSERMAN A. Synthetic data for text localisation in natural images[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2315 - 2324.
- [25] WANG Z Y, YANG J C, JIN H L, et al. DeepFont: identify your font from an image[C]//Proceedings of the 23rd ACM International Conference on Multimedia, 2015: 451 - 459.
- [26] ZHENG Z D, ZHENG L, YANG Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro [C]//Proceedings of IEEE International Conference on Computer Vision, 2017: 3774 - 3782.
- [27] ROGEZ G, SCHMID C. MoCap-guided data augmentation for 3D pose estimation in the wild [C]//Proceedings of 29th International Conference on Neural Information Processing Systems, 2016: 3108 - 3116.
- [28] GAIDON A, WANG Q, CABON Y, et al. Virtual worlds as proxy for multi-object tracking analysis [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4340 - 4349.
- [29] PARMAR N, VASWANI A, USZKOREIT J, et al. Image transformer [C]//Proceedings of the 35th International Conference on Machine Learning, 2018: 4055 - 4064.
- [30] KINGMA D P, BA J A. A method for stochastic optimization[C]//Proceedings of International Conference on Learning Representations, 2015.
- [31] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [C]//Proceedings of International Conference on Learning Representations, 2018.
- [32] 杜兰, 刘彬, 王燕, 等. 基于卷积神经网络的SAR图像目标检测算法[J]. 电子与信息学报, 2016, 38(12): 3018 - 3025.
- DU L, LIU B, WANG Y, et al. Target detection method based on convolutional neural network for SAR image[J]. Journal of Electronics & Information Technology, 2016, 38(12): 3018 - 3025. (in Chinese)
- [33] SALIMANS T, GOODFELLOW I J, ZAREMBA W, et al. Improved techniques for training GANs [C]//Proceedings of 30th International Conference on Neural Information Processing Systems, 2016: 2234 - 2242.
- [34] LEE D H. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks [C]//Proceedings of International Conference on Machine Learning, 2013.
- [35] ABADI M, AGARWAL A, BARHAM P, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems[EB/OL]. (2016 - 03 - 16) [2020 - 08 - 20]. <https://arxiv.org/abs/1603.04467>.
- [36] LI J W, QU C W, SHAO J Q. Ship detection in SAR images based on an improved faster R-CNN [C]//Proceedings of SAR in Big Data Era: Models, Methods and Applications, 2017: 1 - 6.
- [37] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium [C]//Proceedings of 31st Conference on Neural Information Processing Systems, 2017: 6629 - 6640.
- [38] XU Q T, HUANG G, YUAN Y, et al. An empirical study on evaluation metrics of generative adversarial networks [EB/OL]. (2018 - 08 - 17) [2020 - 08 - 20]. <https://arxiv.org/abs/1806.07755v2>.
- [39] HE H, WANG S C, YANG D F, et al. SAR target recognition and unsupervised detection based on convolutional neural network [C]//Proceedings of Chinese Automation Congress, 2017: 435 - 438.
- [40] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of the 14th European Conference on Computer Vision, 2016.
- [41] 李广帅, 苏娟, 李义红. 基于改进 Faster R-CNN 的 SAR 图像飞机检测算法[J]. 北京航空航天大学学报, 2021, 47(1): 159 - 168.
- LIG S, SU J, LI Y H. An aircraft detection algorithm in SAR image based on improved Faster R-CNN [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(1): 159 - 168. (in Chinese)
- [42] 孙显, 王智睿, 孙元睿, 等. AIR-SARShip - 1.0: 高分辨率 SAR 舰船检测数据集[J]. 雷达学报, 2019, 8(6): 852 - 862.
- SUN X, WANG Z R, SUN Y R, et al. AIR-SARShip - 1.0: high-resolution SAR ship detection dataset [J]. Journal of Radars, 2019, 8(6): 852 - 862. (in Chinese)